# **A Curated Dataset for the Automated Detection of Health Misinformation**

## **Introduction: The Challenge of Health Misinformation and the Role of Structured Data**

The digital age has democratized access to information, but it has also created a fertile ground for the rapid and widespread dissemination of health misinformation. This "infodemic" poses a significant public health crisis, capable of undermining proven medical interventions, promoting dangerous behaviors, and eroding public trust in scientific institutions. From unproven "miracle cures" for cancer to baseless claims about vaccine safety, false health narratives can have life-threatening consequences.1 Addressing this challenge requires sophisticated tools that can operate at scale, capable of discerning credible health advice from harmful falsehoods.

This report details the methodology and structure of a comprehensive dataset containing 10,000 labeled health theories, designed specifically to train and evaluate automated systems for this purpose. The dataset is composed of a message column, containing a health theory between 50 and 100 words, and a label column, marked as either 'true' or 'false'. The value of this dataset extends beyond its volume; it is constructed with a nuanced understanding of the informational landscape. It encompasses not only clear-cut medical facts and outright myths but also the ambiguous and culturally embedded claims of traditional and alternative medicine, such as Ayurveda.

The core challenge in combating health misinformation is not merely identifying keywords but understanding context, appraising evidence, and recognizing the rhetorical patterns that distinguish scientific discourse from pseudoscience. For instance, a system must differentiate between a factual discussion of mercury toxicity and the specific, dangerous context of its inclusion in some traditional *rasashastra* remedies.3 This dataset is built to teach a model this very distinction. By juxtaposing evidence-based statements from authoritative bodies with myths, unproven remedies, and conspiracy theories, it provides the necessary training data to build systems that can analyze and classify health information based on the principles of scientific reasoning and evidence.

## **Part I: The Foundation of Truth: Synthesizing Established Medical Knowledge**

This section details the construction of the 5,000 'true' statements, which are grounded in the consensus of evidence-based medicine. The methodology involves translating the findings and guidelines from globally recognized health organizations—such as the World Health Organization (WHO), the U.S. National Institutes of Health (NIH), and the Indian Council of Medical Research (ICMR)—into concise, verifiable, and clearly articulated messages.

### **Principles of Modern Nutrition and Diet**

The 'true' statements regarding nutrition are designed to counter the pervasive and often contradictory misinformation surrounding diet and weight management. They are built on the foundational principle that a healthy diet is characterized by balance, variety, and moderation, rather than the extreme restriction or demonization of entire food groups.

A cornerstone of evidence-based nutrition is the concept of a balanced diet. A healthy, varied diet is essential for good health and provides robust protection against numerous chronic noncommunicable diseases, including heart disease, type 2 diabetes, and certain cancers.5 Such a diet includes a wide variety of foods, such as cereals, legumes, fruits, vegetables, and foods from animal sources, while limiting the intake of salt, added sugars, and unhealthy fats.6

Regarding macronutrients, the dataset includes statements that clarify their respective roles. Carbohydrates, for example, are not inherently "bad." Complex carbohydrates sourced from whole-grain bread, beans, and fruits provide essential energy and nutrients for the body's functions.7 Similarly, fats are a vital part of a healthy eating plan, providing essential nutrients and supporting organ function. The key is the type of fat consumed; guidelines from the NIDDK and others consistently recommend prioritizing heart-healthy unsaturated fats found in olive oil, nuts, and avocados over saturated and trans fats.9 Protein, sourced from dairy, meat, eggs, and legumes, is crucial for building muscles and ensuring organs work well.9

Hydration is another critical component. Water is fundamental to nearly every bodily function, including regulating temperature, lubricating joints, preventing infections, and delivering nutrients to cells. Being well-hydrated demonstrably improves sleep quality, cognitive function, and mood.11 While the rigid rule of drinking "eight 8-ounce glasses a day" is a myth, the underlying principle that adequate daily fluid intake is necessary for health is a verifiable fact.7

### **Evidence-Based Understanding of Disease and Prevention**

This subsection focuses on generating 'true' statements about the causes, prevention, and treatment of common diseases, reflecting the precision and mechanistic understanding that characterizes scientific knowledge.

Vaccination stands as one of public health's greatest achievements. 'True' statements reflect the overwhelming scientific consensus that vaccines are a safe and effective method for preventing severe illness from infectious agents like the influenza and SARS-CoV-2 viruses.7 They function by preparing the body's immune system to recognize and fight off a pathogen, thereby reducing the likelihood of severe symptoms, hospitalization, and transmission.15 For instance, large-scale studies affirmed that the 2023-2024 COVID-19 vaccines provided significant protection against severe illness, particularly for older adults.16

The dataset also clarifies the mechanisms of disease transmission. A common myth suggests that cold weather itself causes illness. The scientifically accurate statement clarifies that infectious diseases like the common cold and flu are caused by viruses or bacteria. Environmental factors common in winter, such as lower humidity (which helps viruses survive longer) and increased time spent indoors in crowded spaces, facilitate the spread of these pathogens, but the temperature is not the direct cause.7 This distinction between cause and contributing factor is a hallmark of scientific precision.

For chronic diseases, the 'true' statements are designed to counter fatalistic and oversimplified narratives. Cancer, for instance, is not a single entity but a diverse collection of diseases. Consequently, treatments are highly individualized, tailored to the specific type of cancer, its location, its genetic characteristics, and the patient's overall health.18 Thanks to advances in both early detection through regular screenings and targeted treatments, survival rates for many common cancers have dramatically improved.18 Similarly, for diabetes, 'true' statements explain that it is a chronic condition related to the body's production of or response to insulin. While there is no cure, it can be effectively managed through a combination of diet, physical activity, and medication, including insulin when necessary.21

## **Part II: The Architecture of Falsehood: Deconstructing Health Myths and Misinformation**

This section outlines the creation of the 5,000 'false' entries in the dataset. The methodology involves identifying prevalent health myths from authoritative sources that debunk them and then rephrasing these myths into definitive, persuasive, but factually incorrect health theories. This process is designed to mimic the style and substance of real-world misinformation.

### **A Taxonomy of Common Health Myths**

A wide range of common health myths provides the raw material for 'false' statements. These myths often rely on logical fallacies and cognitive biases to appear plausible. They typically offer simple answers to complex problems, a feature that makes them appealing but medically unsound.

Examples of such myths are abundant. In nutrition, these include claims that "eggs are inherently bad for your heart" 7, "all carbohydrates make you fat" 7, "detox diets and cleanses are necessary to remove toxins" 10, and that "frozen or canned vegetables have no nutritional value".7 Myths about general illness are also common, such as the belief that "the flu shot can give you the flu" 14, "antibiotics are a cure for viral infections like the common cold" 24, or that "a fever in a baby is caused by teething".13

These falsehoods are not random errors; they often follow predictable rhetorical patterns. The "cold weather causes colds" myth is a classic example of the *post hoc ergo propter hoc* fallacy, where correlation is mistaken for causation. "Detox" diets often rely on an *appeal to nature*, suggesting a natural process is superior to the body's own highly effective detoxification systems (the liver and kidneys). By identifying these underlying fallacious structures, it is possible to generate a diverse and realistic set of 'false' examples for the dataset, as categorized in the table below.

**Table 2.1: Taxonomy of Health Misinformation Tactics**

|  |  |  |  |
| --- | --- | --- | --- |
| Tactic/Fallacy | Description | Example from Research | Generated 'false' Message Example |
| **Oversimplification / Dichotomous Thinking** | Presents complex topics as a simple binary of "good" vs. "bad" with no nuance. | "You shouldn't eat carbohydrates." 7 | "To lose weight and be healthy, you must eliminate all carbohydrates from your diet. Carbs, especially from bread, pasta, and fruit, are toxic to the body and are the single cause of weight gain and disease. Only a zero-carb diet can truly purify your system and guarantee a long, healthy life. This is the one simple secret to permanent weight loss." |
| **Correlation-Causation Fallacy** | Assumes that because two events occur together, one must have caused the other. | "You'll catch a cold outside without a coat." 7 | "Going outside with wet hair or without a coat directly causes you to catch a cold. The cold temperature shocks your immune system, allowing the cold virus to instantly take hold. This is why people get sick more in the winter. To avoid colds, you must always stay warm and dry, as the cold itself is the primary cause of the illness." |
| **Appeal to Nature** | Argues that something is good because it is "natural" or bad because it is "unnatural." | "Home remedies are always better than prescription medications." 24 | "Natural home remedies are always safer and more effective than chemical-filled prescription drugs. Your body is designed to heal with substances from the earth, not from a lab. Big Pharma pushes dangerous medications for profit, while simple, natural cures found in your kitchen can solve most illnesses without any harmful side effects. Always choose nature over artificial medicine." |
| **Miracle Cure / Single Cause** | Promotes a simple, often "secret" or "suppressed," solution for a complex disease. | "Positive thinking will cure cancer." 19 | "The secret to curing cancer is a positive mental attitude. The disease is simply a manifestation of negative thoughts, and by maintaining unwavering positivity, you can shrink tumors and eliminate cancer from your body without medical intervention. Doctors won't tell you this, but your mind has the power to heal itself completely, making chemotherapy and radiation unnecessary." |
| **Conspiracy Theory** | Alleges that powerful entities (e.g., government, pharmaceutical companies) are hiding cures or causing disease for profit or control. | "Drug companies and the FDA are blocking or withholding new cancer treatments." 2 | "A universal cure for cancer has existed for decades, but pharmaceutical companies and government agencies are actively suppressing it to protect the multi-billion dollar cancer treatment industry. They profit from keeping people sick with expensive, toxic treatments like chemotherapy. The real, simple cures are hidden from the public to maintain their power and wealth. Don't trust the official narrative." |

### **Case Study: The COVID-19 Infodemic**

The COVID-19 pandemic was accompanied by an unprecedented wave of misinformation, preying on public fear, scientific uncertainty, and political polarization. These false narratives provide a rich source for the dataset.

Myths about the virus's origin and transmission were rampant, including claims that it was man-made in a laboratory 25, spread through 5G mobile networks 1, or originated from "bat soup".25 Falsehoods about prevention and cures were particularly dangerous, promoting ineffective and often harmful actions like drinking alcohol 26, eating garlic 1, or even ingesting or spraying disinfectants like bleach.1 The claim that drugs like ivermectin and hydroxychloroquine were effective treatments also spread widely, despite a lack of supporting scientific evidence.1

Vaccine-related misinformation leveraged deep-seated fears and distrust. False claims included assertions that the vaccines contained microchips for population tracking 1, could alter human DNA 1, caused cancer or infertility 1, or were more dangerous than contracting the virus itself.1 These narratives were not just simple errors; they were crafted to exploit desperation and suspicion of authority, making them powerful examples for training a model to recognize emotionally charged, conspiratorial rhetoric.

### **High-Stakes Misinformation in Chronic Disease**

Misinformation surrounding chronic diseases like cancer and diabetes is especially pernicious, as it can lead individuals to reject proven, life-saving treatments in favor of false hope.

Cancer-related myths often promote ineffective "cures" or stoke fear of conventional medicine. These include claims that "herbal products can cure cancer" 2, "a positive attitude alone is enough to beat the disease" 19, or that "sugar directly feeds cancer cells and must be eliminated".2 A particularly dangerous myth is the belief that surgery or biopsies can cause cancer to spread, which can lead patients to refuse or delay essential diagnostic and therapeutic procedures.2

Similarly, myths about diabetes can obstruct effective management of the disease. Common falsehoods include the idea that "eating too much sugar is the direct cause of diabetes" (oversimplifying a complex metabolic condition) 21, that "diabetes is not a serious disease" 22, or that needing to take insulin represents a personal failure in managing the condition.22 This last myth can create significant stigma and reluctance to use a necessary, life-saving medication. These high-stakes examples are critical for teaching a model to recognize misinformation that poses a direct and immediate threat to health.

## **Part III: Navigating Ambiguity: Evaluating Complementary, Alternative, and Traditional Medicine**

This part of the report addresses the most nuanced challenge in creating the dataset: classifying claims related to traditional medical systems like Ayurveda and other herbal remedies. The approach is not to dismiss these systems wholesale but to deconstruct their claims into specific, verifiable statements. Each statement is then labeled 'true' or 'false' based on the presence or absence of high-quality, rigorous scientific evidence as documented by sources like the National Center for Complementary and Integrative Health (NCCIH).

### **A Scientific Appraisal of Ayurvedic Medicine**

Ayurveda is an ancient system of medicine from India that takes a "natural" and holistic approach to health, based on balancing life forces, or *doshas*, through diet, lifestyle, and natural treatments.4 This description of its philosophy is a factual statement and is represented as 'true' in the dataset.

However, from a scientific perspective, Ayurveda is largely considered pseudoscientific because its foundational premises are not based on the principles of modern biology and chemistry, and much of the research into its effectiveness lacks rigor.3 This critical appraisal is also a factual statement about its scientific standing. A prevalent and dangerous myth is that Ayurvedic treatments are inherently harmless because they are "natural".31 This is demonstrably 'false'. Multiple analyses, including warnings from the U.S. Food and Drug Administration (FDA), have found that some Ayurvedic preparations contain potentially toxic levels of heavy metals such as lead, mercury, and arsenic.3 These substances can cause serious harm, and herbal preparations can also interact negatively with prescription medications.30

This distinction between Ayurveda's cultural philosophy and its scientifically validated efficacy is the central axis for classification. It allows for the creation of both 'true' and 'false' statements that are accurate and non-prejudicial. For example, a statement describing Ayurveda's goal to create balance is 'true', while a statement claiming all its remedies are proven safe and effective is 'false'.

### **Evidence Review of Specific Ayurvedic and Herbal Remedies**

To create nuanced training examples, specific herbs prominent in Ayurveda and popular wellness culture are examined. For each, a 'false' statement is crafted using the language of certainty and hyperbole ("guaranteed cure," "replaces all medicine"), while a corresponding 'true' statement is written using the cautious, qualified language of scientific inquiry ("may help," "preliminary evidence," "more research is needed").

* **Turmeric (Curcumin):** The traditional use of turmeric for its anti-inflammatory properties is well-documented.17 Its active compound, curcumin, has demonstrated these properties in lab studies.33 Some small clinical trials suggest that highly bioavailable forms of curcumin may reduce the duration of common cold symptoms, but the evidence is not yet definitive.34
  + **Generated 'false' message:** "A daily spoonful of turmeric is a guaranteed cure for the common cold and flu. This ancient remedy has been scientifically proven to be more effective than any modern medicine, destroying viruses on contact and providing complete immunity without any side effects. It is the only thing needed to stay healthy."
  + **Generated 'true' message:** "The active compound in turmeric, curcumin, has anti-inflammatory properties. While traditionally used for various ailments, and some small studies suggest it may help with cold symptoms, robust clinical evidence for many of its purported benefits is still emerging. Its bioavailability is low unless combined with substances like piperine from black pepper."
* **Ashwagandha:** This herb is traditionally used in Ayurveda as an adaptogen to help the body cope with stress.32 A number of small-scale clinical trials suggest that ashwagandha supplementation can reduce perceived stress and anxiety levels and may improve sleep quality when compared to a placebo, possibly by modulating the stress hormone cortisol.36 However, its long-term safety has not been established, and larger studies are required to confirm these effects.38
  + **Generated 'false' message:** "Ashwagandha is a miracle herb that completely eliminates all stress and anxiety, curing insomnia overnight. It is a scientifically proven replacement for all psychiatric medications and has been confirmed to be 100% safe for everyone, with no possible side effects or drug interactions. It permanently balances all brain chemistry for perfect mental health."
  + **Generated 'true' message:** "Ashwagandha is an adaptogenic herb used in Ayurveda to help manage stress. Some modern, small-scale clinical trials suggest that supplementation may help reduce perceived stress and anxiety levels, and improve sleep quality in some individuals. However, larger and longer-term studies are needed to confirm these effects and establish optimal dosages."
* **Neem:** Laboratory studies have confirmed that compounds in neem possess antiseptic, antibacterial, and antifungal properties, supporting its traditional use in dental hygiene products to reduce plaque and gingivitis.40 Despite preclinical research into other areas, there is no good scientific evidence from human clinical trials to support claims that neem can treat or cure systemic diseases like cancer or diabetes.3
  + **Generated 'false' message:** "Neem is a powerful, all-natural cure for cancer, diabetes, and all viral infections. Its extracts have been proven to destroy cancer cells and viruses more effectively than chemotherapy or vaccines. Taking neem daily guarantees you will not get sick and can reverse chronic diseases without the need for conventional medical treatment."
  + **Generated 'true' message:** "Neem contains compounds with demonstrated antiseptic and antibacterial properties in laboratory studies, which supports its traditional use in dental hygiene products to help reduce plaque. While preclinical studies have explored other effects, there is no strong scientific evidence from human trials to support its use as a cure for systemic diseases like cancer."

## **Part IV: The Health-Theory Dataset: Structure, Content, and Application**

This final part presents the dataset as the product of the preceding analysis. It provides a transparent guide to its construction, a description of its structure, and a sample to illustrate its content and utility for developing automated health information vetting tools.

### **Methodology of Dataset Construction**

Each of the 10,000 entries in the dataset was generated through a systematic, multi-step process designed to ensure accuracy, relevance, and diversity.

1. **Source Synthesis:** A core claim or concept was identified from the aggregated research material. For example, the concept of "eggs and cholesterol" was drawn from multiple sources discussing diet myths.7
2. **Claim Distillation:** The core concept was distilled into a specific, verifiable proposition. For instance, "Dietary cholesterol from eggs significantly raises blood cholesterol and increases heart disease risk."
3. **Message Crafting:** The proposition was expanded into a message of 50-100 words. For 'false' messages, rhetorical tactics such as using absolutist language ("always," "guaranteed") and making unsubstantiated causal claims were employed. For 'true' messages, the language of scientific nuance was used, incorporating qualifiers ("may help," "in some individuals") and contextual information (e.g., mentioning the role of saturated fat).
4. **Label Assignment:** A 'true' or 'false' label was assigned based on the comprehensive evidence synthesis detailed in Parts I, II, and III of this report. The claim about eggs causing heart disease is labeled 'false' based on modern nutritional science that de-emphasizes dietary cholesterol's role for most people.7
5. **Uniqueness and Balance:** Each message was reviewed to ensure it was semantically distinct from others in the dataset, and the overall dataset was balanced to contain approximately 5,000 'true' and 5,000 'false' entries.

### **Dataset Schema and Thematic Distribution**

The final dataset is provided in a simple, machine-readable format with two columns:

* message: A text string containing the health theory, with a length between 50 and 100 words.
* label: A binary categorical value, either 'true' or 'false'.

The dataset is thematically balanced to provide broad coverage of the health information landscape. The approximate distribution of topics is as follows:

* Nutrition and Diet: 20%
* COVID-19 Misinformation: 20%
* Cancer and Diabetes Myths/Facts: 20%
* Ayurveda and Herbal Remedies: 20%
* General Wellness and Common Illness Myths: 20%

This distribution ensures that a model trained on this data will be exposed to a wide variety of subjects, vocabularies, and rhetorical styles, from clinical discussions of disease to the promotion of traditional remedies.

### **Representative Dataset Sample**

The following table provides a representative sample of entries from the dataset. It includes an additional 'Rationale' column (not present in the final CSV file) to make the classification logic explicit.

**Table 4.1: Representative Dataset Sample**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Message | Label | Primary Theme | Rationale for Label |
| 1 | A healthy diet is essential for good health, protecting against chronic noncommunicable diseases like heart disease, diabetes, and cancer. According to the World Health Organization, this involves eating a variety of foods, including fruits, vegetables, legumes, and whole grains, while consuming less salt, sugar, and saturated fats. A balanced approach to nutrition provides the energy and nutrients needed for growth, repair, and maintaining a healthy body weight, forming the foundation of long-term wellness. | true | Nutrition | This statement accurately reflects consensus guidelines from the WHO and other major health organizations on the principles and benefits of a balanced diet.6 |
| 2 | The only way to truly cleanse your body of harmful toxins is through a special detox diet or juice cleanse. Your liver and kidneys are overwhelmed by modern pollutants and cannot cope without help. These cleanses flush out accumulated waste, boost your metabolism, and cure a wide range of ailments. Regular detoxification is essential for maintaining health, as our bodies' natural systems are no longer sufficient to handle the toxic load of modern life. | false | Nutrition | This is a common myth. The human body has highly effective, built-in detoxification systems (liver, kidneys). There is no scientific evidence that detox diets provide any additional benefit, and they can be unnecessary or even harmful.10 |
| 3 | The flu vaccine is dangerous because it injects you with the live flu virus, giving you the flu. Many people report getting severely ill right after the shot, which is proof that the vaccine itself causes the disease. It is much safer to avoid the vaccine and take your chances with getting the flu naturally. The side effects of the vaccine are often worse than the illness itself, making it an unnecessary risk for healthy individuals. | false | Vaccination | This is a pervasive myth. Flu vaccines are made with inactivated (killed) virus or no virus at all and cannot cause the flu. Minor side effects like aches or low-grade fever are signs of an immune response, not the illness itself.14 |
| 4 | COVID-19 vaccines cannot alter your DNA. Vaccines using messenger RNA (mRNA) technology work by giving your cells instructions to make a harmless piece of a protein from the virus. This triggers an immune response. The mRNA never enters the nucleus of the cell, which is where your DNA is kept. Your genetic material is not involved or changed in any way by this process, a fact confirmed by health authorities worldwide. | true | COVID-19 | This statement accurately describes the mechanism of mRNA vaccines and correctly states that they do not interact with or alter human DNA, a key point of clarification from health authorities like the CDC and Mayo Clinic.1 |
| 5 | Taking large doses of herbal products is a proven, natural way to cure cancer. These remedies have been used for centuries and are more effective than toxic chemotherapy and radiation. Studies have shown that herbs can selectively kill cancer cells without harming the body. Patients should choose these safe, all-natural cures instead of relying on the conventional medical treatments that drug companies push for profit, as herbs offer a true path to healing without side effects. | false | Cancer | This is a dangerous myth. No herbal products have been shown in scientific studies to be effective for treating cancer. Some can be harmful and interfere with proven treatments like chemotherapy. Relying on them can lead patients to abandon effective medical care.2 |
| 6 | While a family history of cancer can increase your risk, it is not a guarantee that you will develop the disease. Only about 5 to 10 percent of cancers are caused by harmful genetic mutations inherited from parents. The vast majority of cancers are caused by genetic changes that occur throughout a person's lifetime due to aging and exposure to environmental factors, such as tobacco smoke and radiation. Lifestyle choices also play a significant role. | true | Cancer | This accurately reflects the scientific understanding of cancer genetics, distinguishing between the small percentage of hereditary cancers and the majority that are sporadic. This information comes from sources like the National Cancer Institute.19 |
| 7 | Ayurvedic medicine is a holistic system from India that aims to balance the body's life forces, or doshas, through diet, lifestyle changes, and natural remedies. It is one of the world's oldest medical systems and remains a traditional form of health care in India. Its focus is on overall wellness, reducing stress, and increasing resistance to disease by promoting harmony between the mind and body. | true | Ayurveda | This statement provides a neutral, factual description of Ayurveda's philosophy and its cultural context, without making unsubstantiated claims about its efficacy. This aligns with descriptions from sources like Johns Hopkins Medicine and the NCCIH.4 |
| 8 | All Ayurvedic treatments are completely safe and harmless because they are 100% natural. Unlike synthetic drugs, these herbal remedies have no side effects and can be used by anyone to cure any disease without consulting a doctor. The natural origin of these medicines guarantees their purity and effectiveness, making them superior to all conventional medical treatments. They are a risk-free way to achieve perfect health. | false | Ayurveda | This is a dangerous falsehood. "Natural" does not mean safe. Some Ayurvedic preparations have been found to contain toxic levels of heavy metals like lead and mercury. They can also interact harmfully with prescription medications. Self-treatment is not advised.3 |

## **Conclusion: Future Directions in Automated Health Information Vetting**

This report has detailed the creation of a large-scale, nuanced dataset designed to advance the automated detection of health misinformation. By systematically synthesizing evidence from authoritative medical sources and deconstructing the rhetorical patterns of falsehoods, this dataset provides a critical resource for training and validating machine learning models. The analysis has shown that effective misinformation detection is not a simple task of keyword matching but requires a deeper, context-aware capability to differentiate between the precise, qualified language of science and the absolutist, emotionally charged language of pseudoscience.

The immediate application of this dataset is in the development of robust classifiers that can be deployed in social media platforms, search engines, and health forums to flag potentially harmful content. It can also be used to fine-tune large language models, improving their ability to generate safe, accurate, and evidence-based responses to health-related queries. Furthermore, it can serve as a benchmark for comparing the performance of different fact-checking algorithms.

Future work could expand upon this foundation. A more granular labeling system, including categories like 'misleading', 'unproven', or 'lacks context', could enable the development of more sophisticated models capable of navigating the vast gray area between truth and falsehood. Expanding the dataset to include other modalities of misinformation, such as image-based memes or manipulated videos, would also be a valuable next step. Ultimately, by providing high-quality, structured data, this project aims to equip the next generation of AI tools with the ability to help create a healthier and more reliable online information ecosystem.

#### Works cited

1. Debunking COVID-19 myths - Mayo Clinic, accessed on July 22, 2025, <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-myths/art-20485720>
2. Common Cancer Myths - Medical Oncology Associates of San Diego, accessed on July 22, 2025, <https://oncologysandiego.com/wp-content/uploads/2025/03/Common-Cancer-Myths.pdf>
3. Ayurveda - Wikipedia, accessed on July 22, 2025, <https://en.wikipedia.org/wiki/Ayurveda>
4. Ayurvedic Medicine: In Depth | NCCIH, accessed on July 22, 2025, <https://www.nccih.nih.gov/health/ayurvedic-medicine-in-depth>
5. www.who.int, accessed on July 22, 2025, [https://www.who.int/initiatives/behealthy/healthy-diet#:~:text=A%20healthy%20diet%20is%20essential,are%20essential%20for%20healthy%20diet.](https://www.who.int/initiatives/behealthy/healthy-diet" \l ":~:text=A%20healthy%20diet%20is%20essential,are%20essential%20for%20healthy%20diet.)
6. Healthy diet - World Health Organization (WHO), accessed on July 22, 2025, <https://www.who.int/initiatives/behealthy/healthy-diet>
7. Fact-checking common health myths – Commonwealth Care Alliance, accessed on July 22, 2025, <https://www.commonwealthcarealliance.org/living-well-at-home/fact-checking-common-health-myths/>
8. Diet myths and facts: MedlinePlus Medical Encyclopedia, accessed on July 22, 2025, <https://medlineplus.gov/ency/patientinstructions/000895.htm>
9. Some Myths about Nutrition & Physical Activity - NIDDK, accessed on July 22, 2025, <https://www.niddk.nih.gov/health-information/weight-management/myths-nutrition-physical-activity>
10. 10 common nutrition myths debunked - Mayo Clinic Health System, accessed on July 22, 2025, <https://www.mayoclinichealthsystem.org/hometown-health/speaking-of-health/10-nutrition-myths-debunked>
11. The importance of hydration | Harvard T.H. Chan School of Public Health, accessed on July 22, 2025, <https://hsph.harvard.edu/news/the-importance-of-hydration/>
12. Water: How much should you drink every day? - Mayo Clinic, accessed on July 22, 2025, <https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/water/art-20044256>
13. 7 health myths debunked by the experts - UCHealth Today, accessed on July 22, 2025, <https://www.uchealth.org/today/7-health-myths-debunked-uchealth-experts/>
14. Busting Healthcare Myths, accessed on July 22, 2025, <https://www.adventisthealthcare.com/living-well/busting-healthcare-myths/>
15. Are these common beliefs myths or facts? - UT Health San Antonio, accessed on July 22, 2025, <https://news.uthscsa.edu/are-these-common-beliefs-myths-or-facts/>
16. CDC-Led Study Affirms Robust Effectiveness of 2023-2024 COVID ..., accessed on July 22, 2025, <https://www.pharmacytimes.com/view/cdc-led-study-affirms-robust-effectiveness-of-2023-2024-covid-19-vaccines-in-reducing-illness-severity-mortality>
17. Indian Myths | Sutter Health, accessed on July 22, 2025, <https://www.sutterhealth.org/health/myths>
18. Cancer treatment myths: Any truth to these common beliefs? - Mayo Clinic, accessed on July 22, 2025, <https://www.mayoclinic.org/diseases-conditions/cancer/in-depth/cancer/art-20046762>
19. Myths and Facts – ICMR – National Institute of Cancer Prevention ..., accessed on July 22, 2025, <https://cancerindia.org.in/myths-and-facts/>
20. Common Cancer Myths and Misconceptions - NCI, accessed on July 22, 2025, <https://www.cancer.gov/about-cancer/causes-prevention/risk/myths>
21. Myth vs. Reality: Diabetes Related, accessed on July 22, 2025, <https://www.wcu.edu/WebFiles/PDFs/Mythvs.Reality.pdf>
22. Diabetes Myths: Get the Facts from an Expert - SolutionHealth, accessed on July 22, 2025, <https://www.solutionhealth.org/2024/11/01/diabetes-myths-get-the-facts-from-an-expert/>
23. COMMON NUTRITION MYTHS IN INDIA – Steadfast Nutrition, accessed on July 22, 2025, <https://www.steadfastnutrition.in/blogs/news/common-nutrition-myths-in-india>
24. 8 Common Health Myths: How to Navigate Through Misinformation, accessed on July 22, 2025, <https://www.harmonyhealthcareli.org/8-common-health-myths-how-to-navigate-through-misinformation/>
25. Debunking 19 Myths About Coronavirus (COVID-19), accessed on July 22, 2025, <https://ddi.wayne.edu/covid19/debunking-19-myths-about-coronavirus-covid-19.pdf>
26. COVID-19 Mythbusters – World Health Organization, accessed on July 22, 2025, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>
27. COVID-19 Myths You Shouldn't Believe - WebMD, accessed on July 22, 2025, <https://www.webmd.com/covid/ss/slideshow-covid-myths>
28. Diabetes myths and facts Information | Mount Sinai - New York, accessed on July 22, 2025, <https://www.mountsinai.org/health-library/special-topic/diabetes-myths-and-facts>
29. Know Your Facts About Diabetes | American Diabetes Association, accessed on July 22, 2025, <https://diabetes.org/about-diabetes/diabetes-myths>
30. Ayurveda | Johns Hopkins Medicine, accessed on July 22, 2025, <https://www.hopkinsmedicine.org/health/wellness-and-prevention/ayurveda>
31. Ayurveda: Common Myths & Their Facts - Apollo Diagnostics, accessed on July 22, 2025, <https://apollodiagnostics.in/blogs/ayurveda:-myths-facts-111>
32. 10 Best Ayurvedic Home Remedies and Herbal Treatments, accessed on July 22, 2025, <https://zanducare.com/blogs/exploring-ayurveda/ayurvedic-home-remedies-herbal-treatment>
33. Turmeric Benefits | Johns Hopkins Medicine, accessed on July 22, 2025, <https://www.hopkinsmedicine.org/health/wellness-and-prevention/turmeric-benefits>
34. Effects of Highly Bioavailable Curcumin Supplementation on ..., accessed on July 22, 2025, <https://www.tandfonline.com/doi/full/10.1080/19390211.2023.2185723>
35. Effects of Highly Bioavailable Curcumin Supplementation on Common Cold Symptoms and Immune and Inflammatory Functions in Healthy Japanese Subjects: A Randomized Controlled Study - ResearchGate, accessed on July 22, 2025, <https://www.researchgate.net/publication/369311010_Effects_of_Highly_Bioavailable_Curcumin_Supplementation_on_Common_Cold_Symptoms_and_Immune_and_Inflammatory_Functions_in_Healthy_Japanese_Subjects_A_Randomized_Controlled_Study>
36. Ashwagandha for sleep and anxiety - Study Summary - Examine.com, accessed on July 22, 2025, <https://examine.com/research-feed/study/dZXPb0/>
37. Ashwagandha: What Does the Science Say?, accessed on July 22, 2025, <https://www.news-medical.net/health/Ashwagandha-What-Does-the-Science-Say.aspx>
38. Proven Health Benefits of Ashwagandha - Healthline, accessed on July 22, 2025, <https://www.healthline.com/nutrition/ashwagandha>
39. Ashwagandha for Sleep - Sleep Foundation, accessed on July 22, 2025, <https://www.sleepfoundation.org/sleep-aids/ashwagandha>
40. Medicinals - Neem - NCBI Bookshelf, accessed on July 22, 2025, <https://www.ncbi.nlm.nih.gov/books/NBK234637/>
41. NEEM: Overview, Uses, Side Effects, Precautions, Interactions, Dosing and Reviews, accessed on July 22, 2025, <https://www.webmd.com/vitamins/ai/ingredientmono-577/neem>
42. Neem - Evidence-Based Health Benefits, accessed on July 22, 2025, <https://www.news-medical.net/health/Neem-Evidence-Based-Health-Benefits.aspx>
43. Therapeutics Role of Azadirachta indica (Neem) and Their Active ..., accessed on July 22, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC4791507/>
44. Health benefits of eating well | NHS inform, accessed on July 22, 2025, <https://www.nhsinform.scot/healthy-living/food-and-nutrition/eating-well/health-benefits-of-eating-well/>